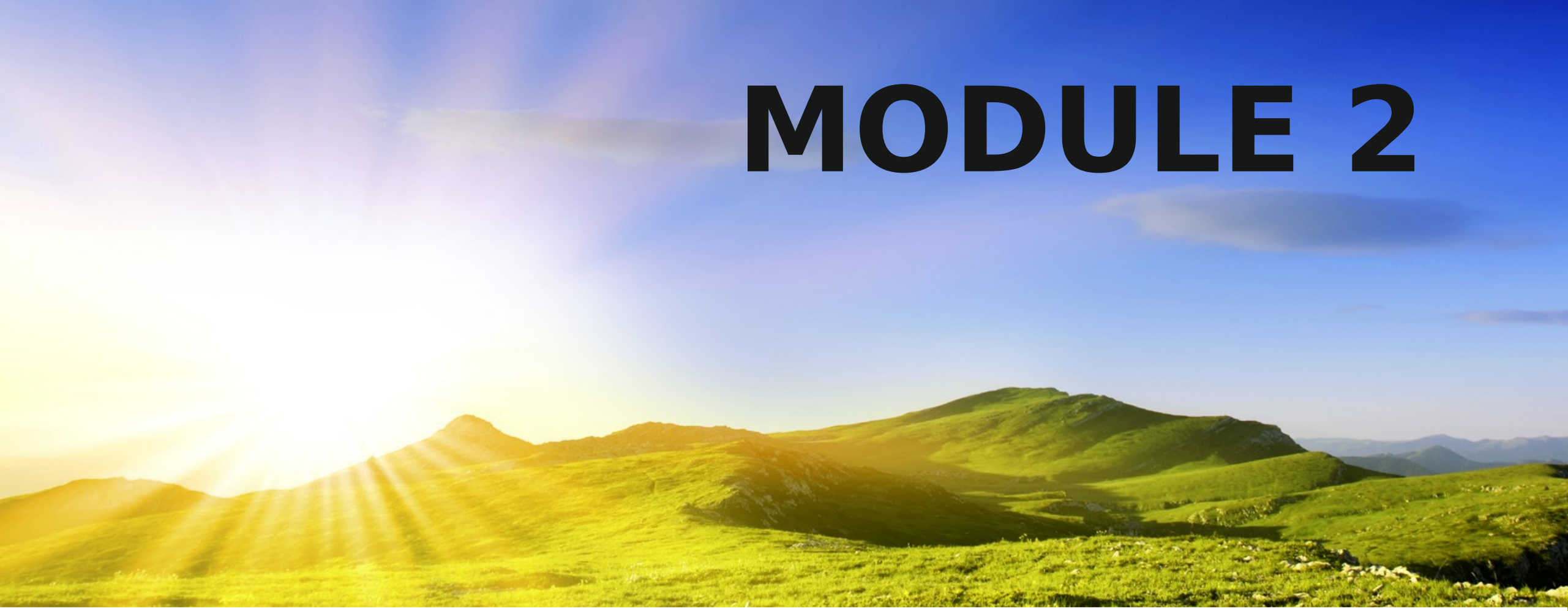


MODULE 2

Frequency Distributions and Graph Presentations

FREQUENCY DISTRIBUTIONS

Frequency Distributions

- After collecting data, the first task for a researcher is to organize and simplify the data so that it is possible to get a general overview of the results.
- This is the goal of descriptive statistical techniques.
- One method for simplifying and organizing data is to construct a frequency distribution.

A **frequency distribution** is the organization of raw data in table form, using classes and frequencies.

The categorical frequency distribution is used for data that can be placed in specific categories

Sample 1

25 people were given a blood test to determine their blood type. Construct a categorical frequency distribution for the data given.

A	B	B	AB	O
O	O	B	AB	B
B	B	O	A	O
A	O	O	O	AB
AB	A	O	B	A

A Class	B Tally	C Frequency	D Percent
A			
B			
O			
AB			

Sample 1

A Class	B Tally	C Frequency	D Percent
A	𝍸	5	20
B	𝍸 //	7	28
O	𝍸 ////	9	36
AB	////	4	16
		Total 25	100

Grouped Frequency Distribution

- When the range of the data is large, the data must be grouped into classes that are more than one unit in width, in what is called a grouped frequency distribution. For example, a distribution of the number of hours that boat batteries lasted is the following:

Class limits	Class boundaries	Tally	Frequency
24–30	23.5–30.5	///	3
31–37	30.5–37.5	/	1
38–44	37.5–44.5	////	5
45–51	44.5–51.5	//// ////	9
52–58	51.5–58.5	//// /	6
59–65	58.5–65.5	/	1
			25

- Students sometimes have difficulty finding class boundaries when given the class limits.
- The basic rule is that *the class limits should have the same decimal place value as the data, but the class boundaries should have one additional place value and end in a 5.*
- For example,

if the values in the data set are whole numbers, such as 24, 32, and 18, the limits for a class

might be 31–37, and the boundaries are 30.5– 37.5. Find the boundaries by subtracting 0.5

from 31 (the lower class limit) and adding 0.5 to 37 (the upper class limit).

$$\text{Lower limit} - 0.5 = 31 - 0.5 = 30.5 = \text{lower boundary}$$
$$\text{Upper limit} + 0.5 = 37 + 0.5 = 37.5 = \text{upper boundary}$$

If the data are in tenths, such as 6.2, 7.8, and 12.6, the limits for a class hypothetically might be 7.8–8.8, and the boundaries for that class would be 7.75– 8.85. Find these values by

- The researcher must decide how many classes to use and the width of each class. To construct a frequency distribution, follow these rules:

1. There should be between 5 and 20 classes. Although there is no hard-and-fast rule for the number of classes contained in a frequency distribution, it is of the utmost importance to have enough classes to present a clear description of the collected data.

2. It is preferable but not absolutely necessary that the class width be an odd number. This ensures that the midpoint of each class has the same place value as the data. The class midpoint Xm is obtained by adding the lower and upper boundaries and dividing by 2, or adding the lower and upper

■ **The researcher must decide how many classes to use and the width of each class. To construct a frequency distribution, follow these rules:**

3. The classes must be mutually exclusive. (not overlapping)

Age
10–20
20–30
30–40
40–50

Age
10–20
21–31
32–42
43–53

4. The classes must be continuous. (No gaps)

5. The classes must be exhaustive. There should be enough classes to accommodate all the data.

- **The researcher must decide how many classes to use and the width of each class. To construct a frequency distribution, follow these rules:**

6. The classes must be equal in width. This avoids a distorted view of the data. One exception occurs when a distribution has a class that is open-ended.

Example of an open-ended distribution:

Age	Frequency	Minutes	Frequency
10–20	3	Below 110	16
21–31	6	110–114	24
32–42	4	115–119	38
43–53	10	120–124	14
54 and above	8	125–129	5

Example:

Record High Temperatures

These data represent the record high temperatures in degrees Fahrenheit (°F) for each of the 50 states. Construct a grouped frequency distribution for the data using 7 classes.

112	100	127	120	134	118	105	110	109	112
110	118	117	116	118	122	114	114	105	109
107	112	114	115	118	117	118	122	106	110
116	108	110	121	113	120	119	111	104	111
120	113	120	117	105	110	118	112	114	114

Source: *The World Almanac and Book of Facts.*

SOLUTION:

The procedure for constructing a grouped frequency distribution for numerical data follows.

Step 1 Determine the classes.

Find the highest value and lowest value: $H = 134$ and $L = 100$.

Find the range: R = highest value — lowest value = $H - L$, so

$R = 134 - 100 = 34$

Select the number of classes desired (usually between 5 and 20). In this case, 7 is arbitrarily chosen.

Find the class width by dividing the range by the number of classes.

$$\text{Width} = \frac{R}{\text{number of classes}} = \frac{34}{7} = 4.9$$

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies.

The completed frequency distribution is

Class limits	Class boundaries	Tally	Frequency
100–104	99.5–104.5	//	2
105–109	104.5–109.5	𝍸 ///	8
110–114	109.5–114.5	𝍸 𝍸 𝍸 ///	18
115–119	114.5–119.5	𝍸 𝍸 ///	13
120–124	119.5–124.5	𝍸 //	7
125–129	124.5–129.5	/	1
130–134	129.5–134.5	/	1
			$n = \sum f = 50$

The frequency distribution shows that the class 109.5–114.5 contains the largest number of temperatures (18) followed by the class 114.5–119.5 with 13 temperatures. Hence, most of the temperatures (31) fall between 109.5 and 119.5°F.

CUMULATIVE FREQUENCY

- **A cumulative frequency distribution is a distribution that shows the number of data values less than or equal to a specific value (usually an upper boundary). The values are found by adding the frequencies of the classes less than or equal to the upper class boundary of a specific class. This gives an ascending cumulative frequency.**

The cumulative frequency distribution for the data in this example is as follows:

Frequency
2
8
18
13
7
1
1
$n = \sum f = 50$

	Cumulative frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

Cumulative frequencies are used to show how many data values are accumulated up to and including a specific class. In Example 2–2, 28 of the total record high temperatures are less than or equal to 114. Forty-eight of the total record high temperatures are less than or equal to 124.

In summary....

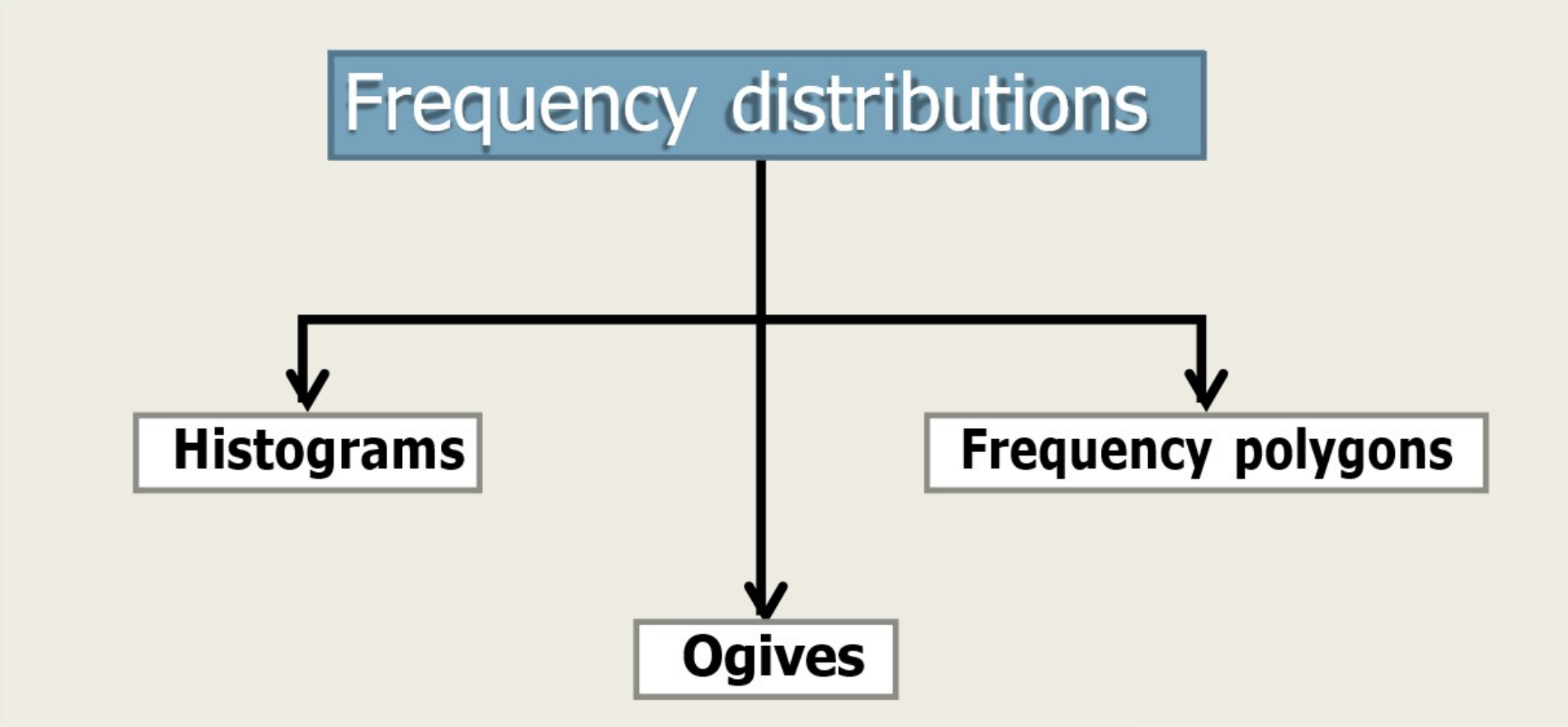
Constructing a Grouped Frequency Distribution

Step 1 Determine the classes.

- Find the highest and lowest values.
- Find the range.
- Select the number of classes desired.
- Find the width by dividing the range by the number of classes and rounding up.
- Select a starting point (usually the lowest value or any convenient number less than the lowest value); add the width to get the lower limits.
- Find the upper-class limits.
- Find the boundaries.

Step 2 Tally the data.

Step 3 Find the numerical frequencies from the tallies, and find the cumulative frequencies.


Frequency distributions
Histograms
Ogives
Frequency polygons

Histograms, Frequency Polygons, and Ogives

- After you have organized the data into a frequency distribution, you can present them in graphical form. The purpose of graphs in statistics is to convey the data to the viewers in pictorial form. It is easier for most people to comprehend the meaning of data presented graphically than data presented numerically in tables or frequency distributions. This is especially true if the users have little or no statistical knowledge.
- Statistical graphs can be used to describe the data set or to analyze it. Graphs are also useful in getting the audience's attention in a publication or a speaking presentation. They can be used to discuss

The Histogram

- The histogram is a graph that displays the data by using contiguous vertical bars (unless the frequency of a class is 0) of various heights to represent the frequencies of the classes.

- Here are the steps:

1. Determine the frequency and relative frequency for each class.
2. Mark the class boundaries on a horizontal measurement axis.
3. Above each class interval, draw a rectangle whose height is the corresponding relative frequency (or frequency).

The Histogram

Example

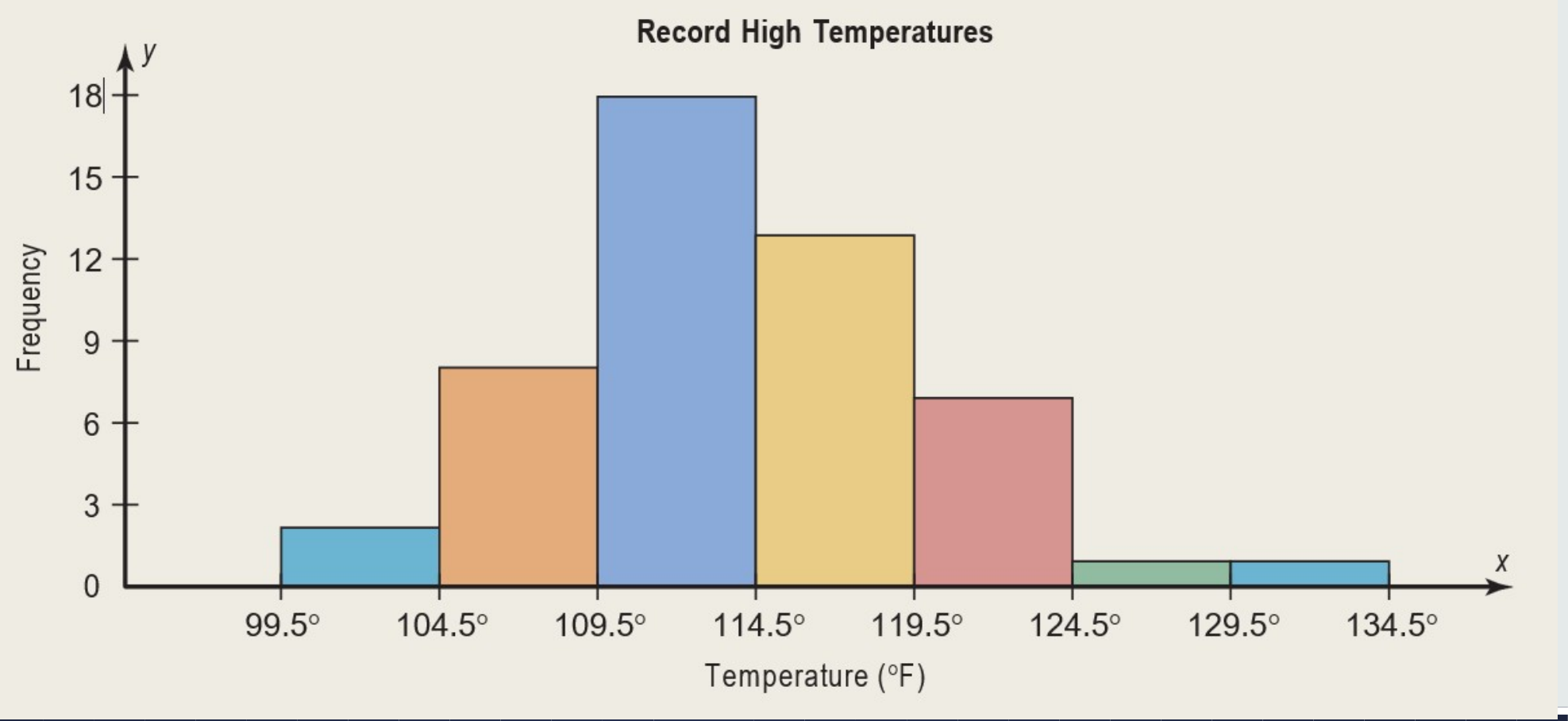
Record High Temperatures

Construct a histogram to represent the data shown for the record high temperatures for each of the 50 states (see Example 2–2).

Class boundaries	Frequency
99.5–104.5	2
104.5–109.5	8
109.5–114.5	18
114.5–119.5	13
119.5–124.5	7
124.5–129.5	1
129.5–134.5	1

The Histogram

Example

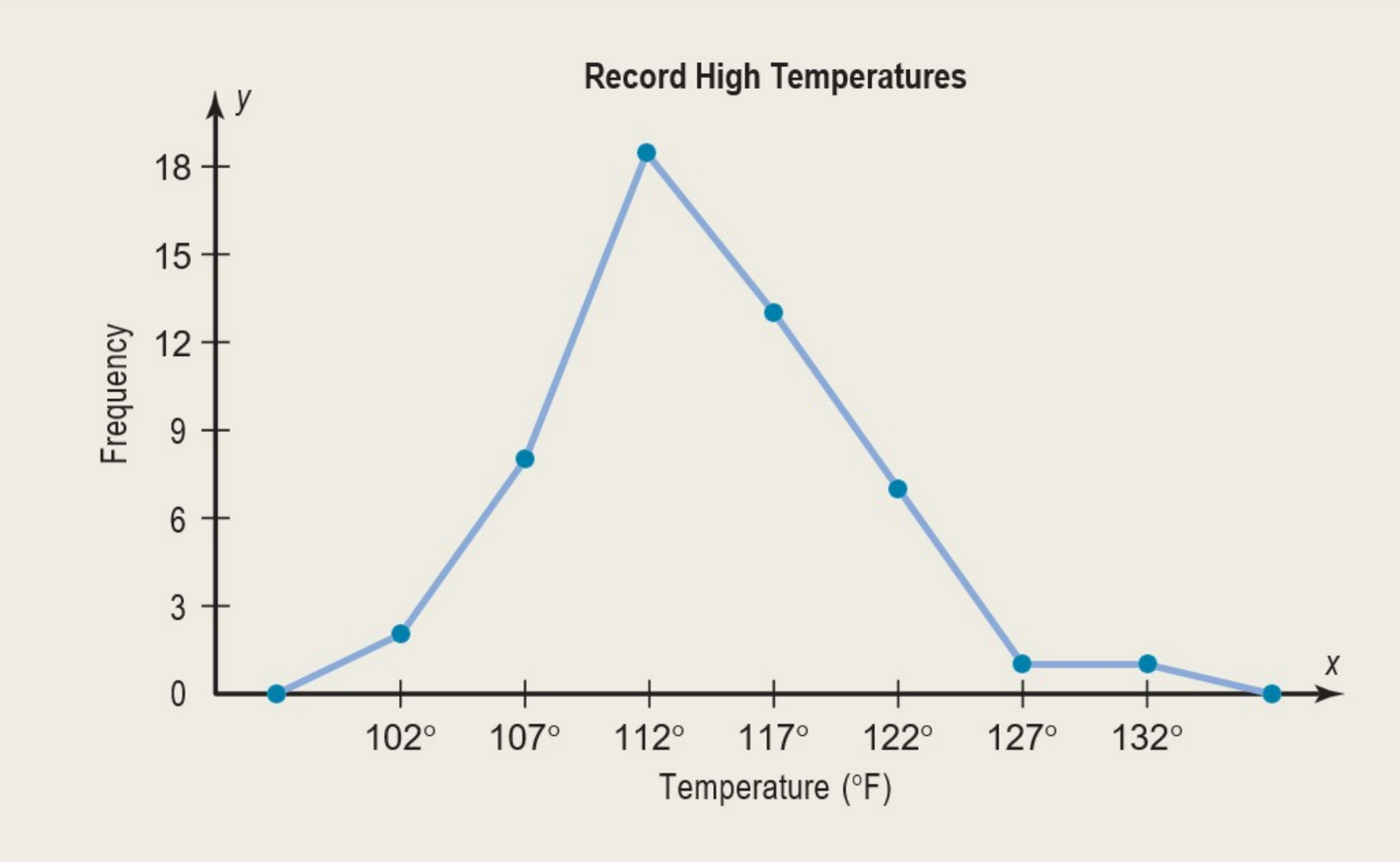


The Frequency Polygon

The *frequency polygon* is a graph that displays the data by using lines that connect points plotted for the frequencies at the midpoints of the classes. The frequencies are represented by the heights of

Class boundaries	Midpoints	Frequency
99.5–104.5	102	2
104.5–109.5	107	8
109.5–114.5	112	18
114.5–119.5	117	13
119.5–124.5	122	7
124.5–129.5	127	1
129.5–134.5	132	1

The Frequency Polygon

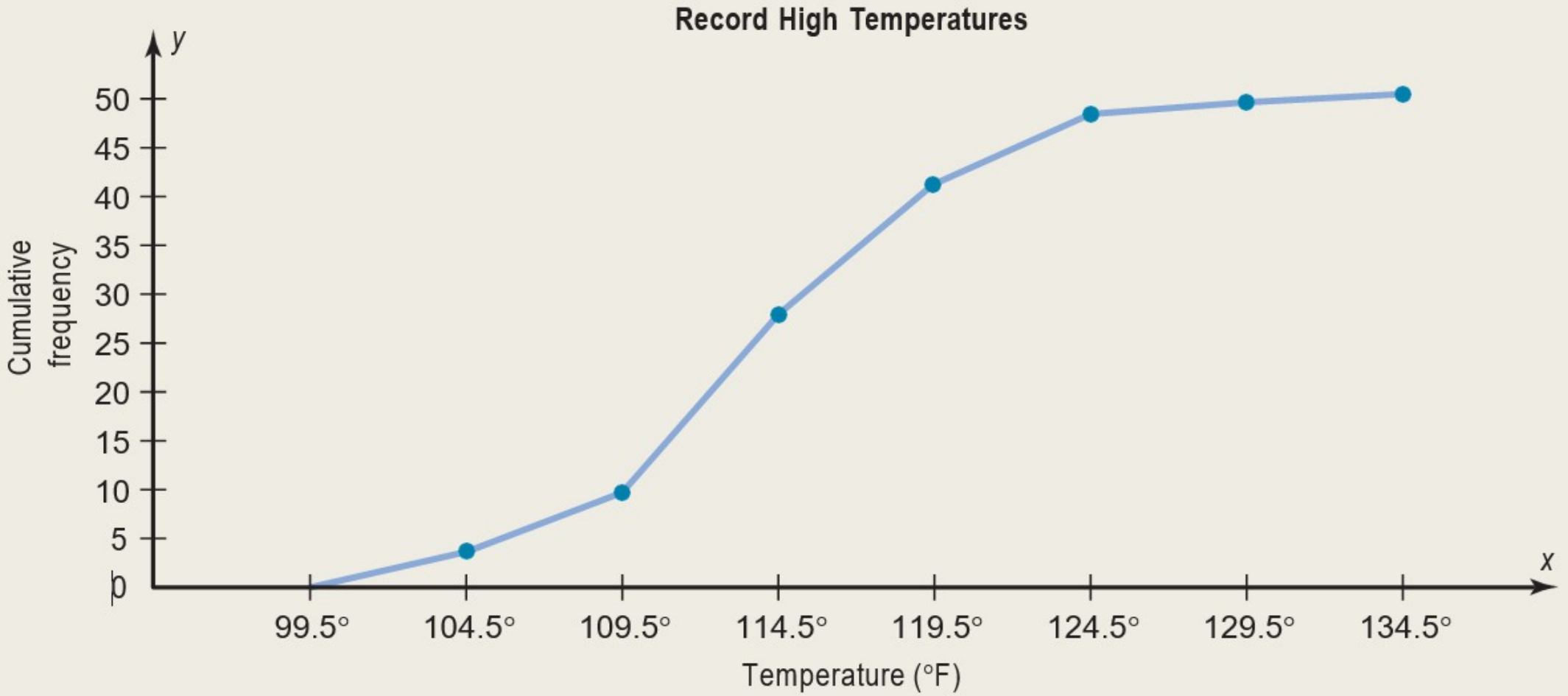


The Ogive

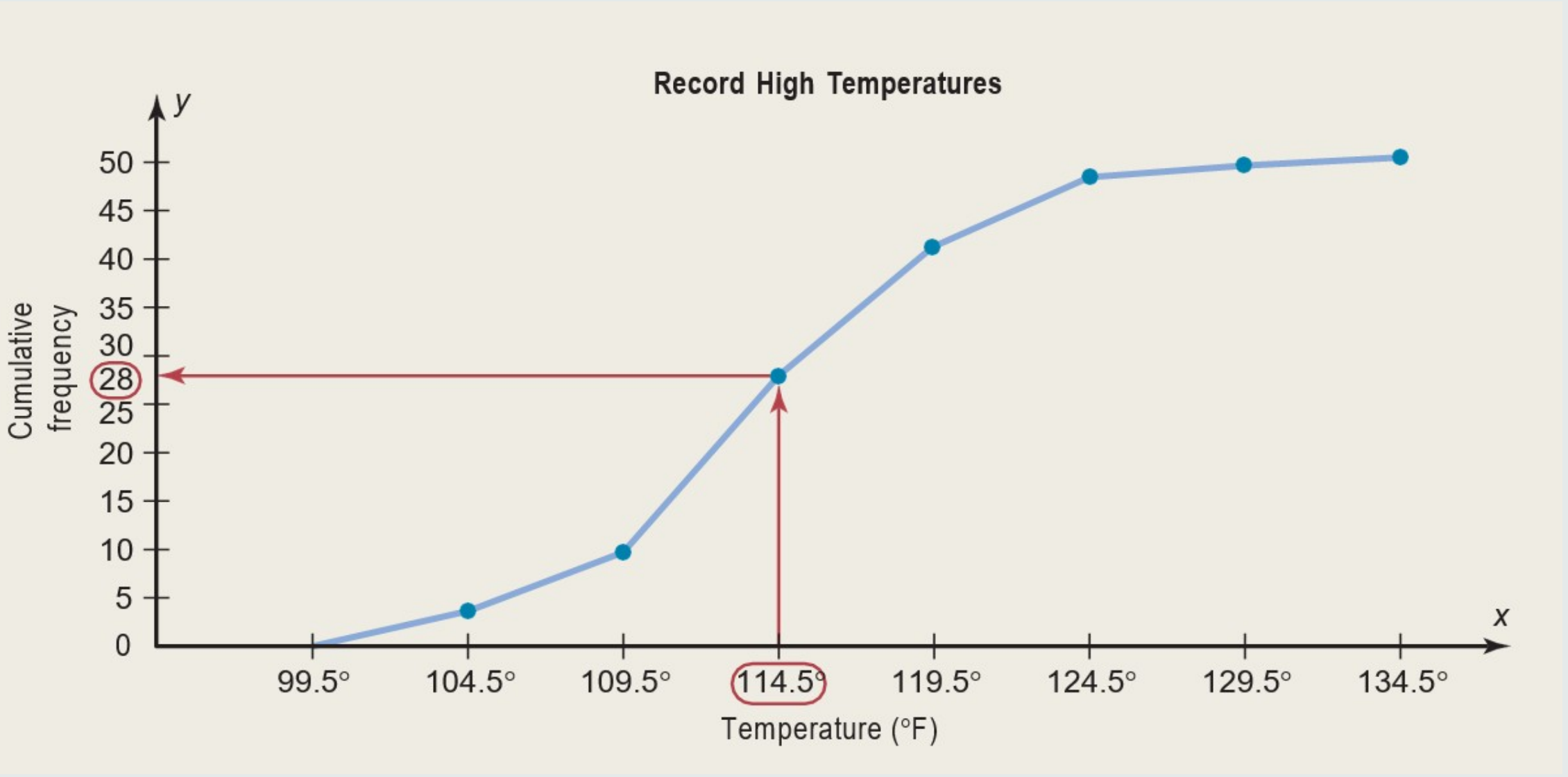
The *ogive* is a graph that represents the cumulative frequencies for the classes in a frequency distribution.

	Cumulative frequency
Less than 99.5	0
Less than 104.5	2
Less than 109.5	10
Less than 114.5	28
Less than 119.5	41
Less than 124.5	48
Less than 129.5	49
Less than 134.5	50

The Ogive


Record High Temperatures
y
50
45
40
35
30
25
20
15
10
5
0
Cumulative frequency
x
99.5°
104.5°
109.5°
114.5°
119.5°
124.5°
129.5°
134.5°
Temperature (°F)

The Ogive


Record High Temperatures
y
50
45
40
35
30
28
25
20
15
10
5
0
Cumulative frequency
x
99.5°
104.5°
109.5°
114.5°
119.5°
124.5°
129.5°
134.5°
Temperature (°F)

In summary….

Constructing Statistical Graphs

Step 1 Draw and label the x and y axes.

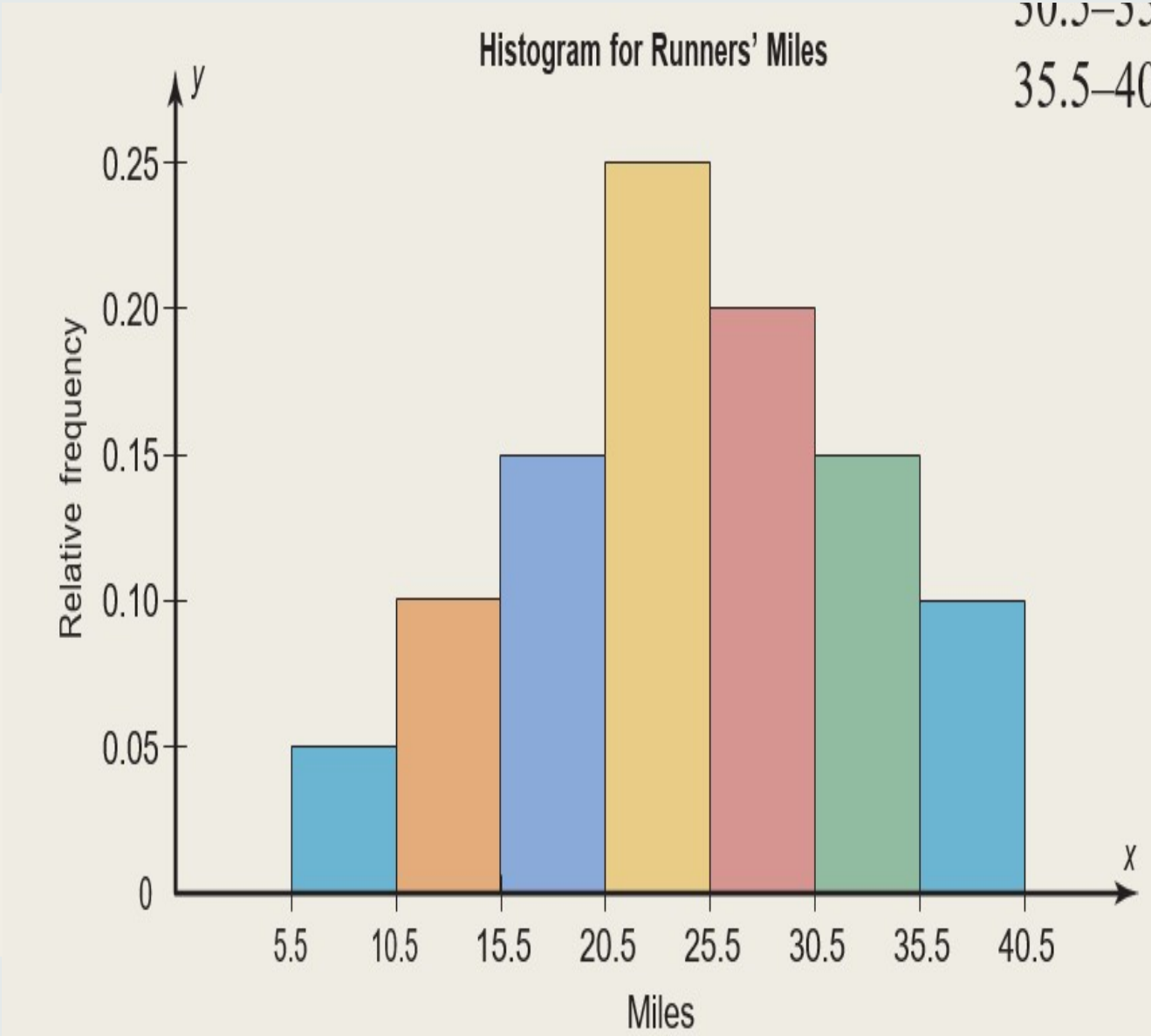
Step 2 Choose a suitable scale for the frequencies or cumulative frequencies, and label it on the y axis.

Step 3 Represent the class boundaries for the histogram or ogive, or the midpoint for the frequency polygon, on the x axis.

Step 4 Plot the points and then draw the bars or lines.

Relative

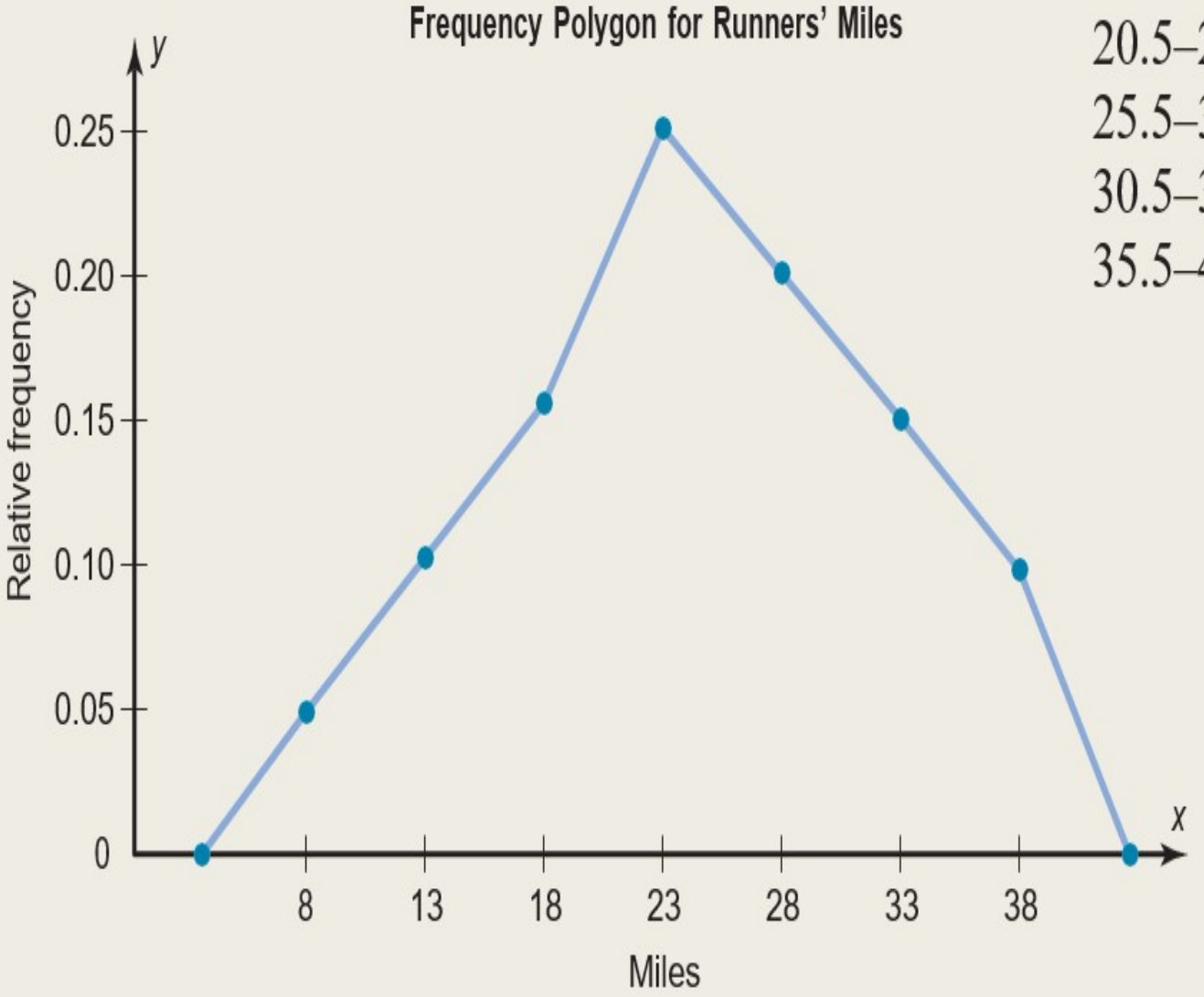
Class boundaries	Frequency	Relative frequency
5.5–10.5	1	0.05
10.5–15.5	2	0.10
15.5–20.5	3	0.15
20.5–25.5	5	0.25
25.5–30.5	4	0.20
30.5–35.5	3	0.15
35.5–40.5	2	0.10
	20	1.00



(a) Histogram

Relative

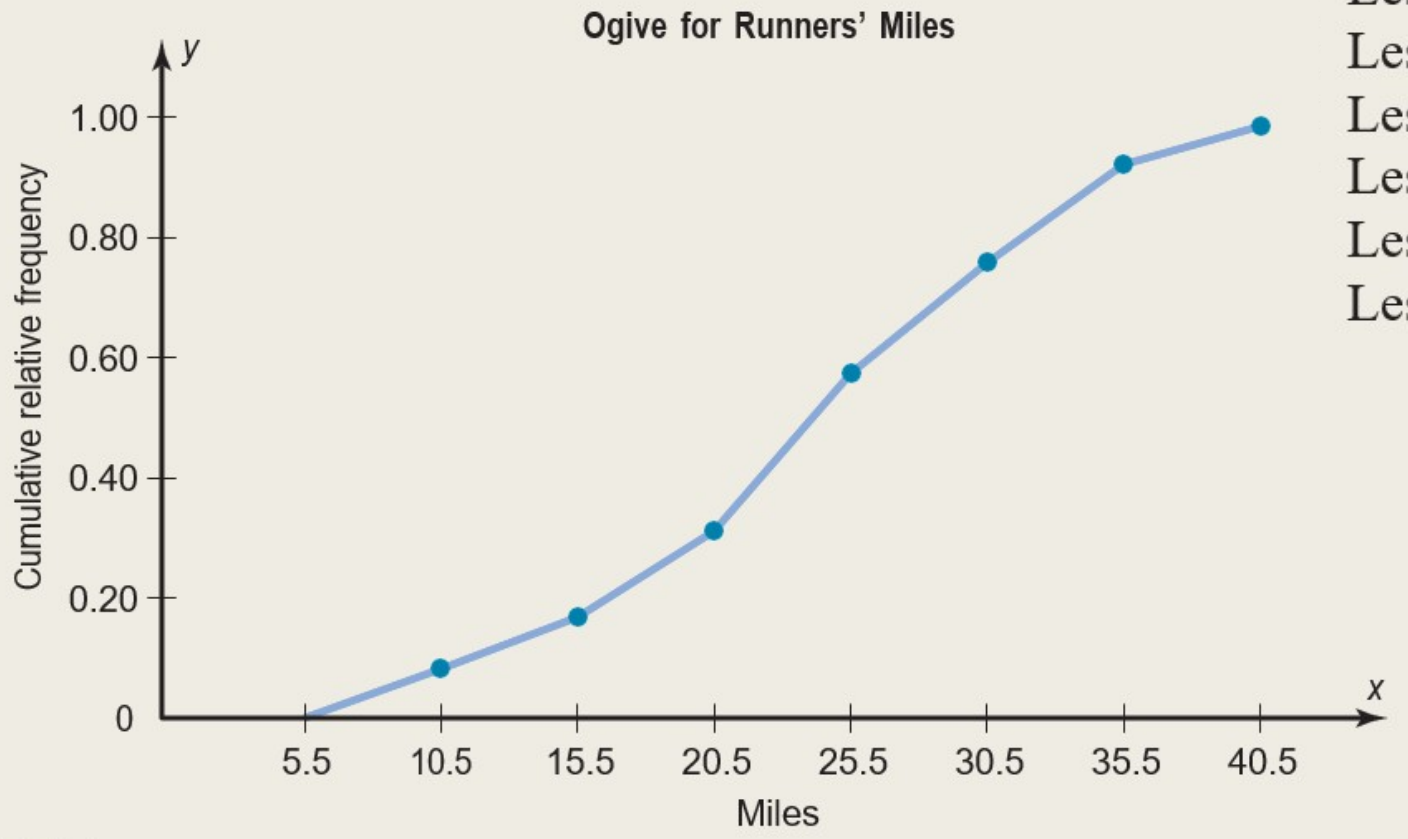
Class boundaries	Midpoints	Relative frequency
5.5–10.5	8	0.05
10.5–15.5	13	0.10
15.5–20.5	18	0.15
20.5–25.5	23	0.25
25.5–30.5	28	0.20
30.5–35.5	33	0.15
35.5–40.5	38	0.10
		1.00



(b) Frequency polygon

15.5–2
20.5–2
25.5–3
30.5–3
35.5–4

Relative Frequency



(c) Ogive

	Cumulative frequency	Cumulative relative frequency
Less than 5.5	0	0.00
Less than 10.5	1	0.05
Less than 15.5	3	0.15
Less than 20.5	6	0.30
Less than 25.5	11	0.55
Less than 30.5	15	0.75
Less than 35.5	18	0.90
Less than 40.5	20	1.00

????
QUESTION ????